ストカスティック言語について (情報科学の数学的理論)

AUTHOR(S):
那須, 正和; 本多, 波雄

CITATION:
那須, 正和 ...[et al]. ストカスティック言語について (情報科学の数学的理論). 数理解析研究所講究録 1971, 123: 91-102

ISSUE DATE:
1971-09

URL:
http://hdl.handle.net/2433/106503

数理解析研究所講究録
123巻 1971年 91-102

ストカスティック言語について

東北大・通研　那須正和　本多波雄

§1　序

有限オートマトンにおいて、初期状態の分布及び状態の遷移が確率的であるようなシステムのモデルを有限確率オートマトンという。すなわち、入力アルファベットΣ上の有限確率オートマトンとは、システム $\mathfrak{A} = \langle S, \pi, \{A(\sigma) \mid \sigma \in \Sigma\}, \eta^F \rangle$ で与えられる。ここでSは有限個(nとする)の状態の集合、πは各要素が非負実数でn個の要素の和が1であるn次元確率行ベクトルであり初期状態の分布を示す。$A(\sigma)$ $\sigma \in \Sigma$ は $n \times n$ 確率遷移行列でその(i, j)要素は状態s_iから状態s_jへ遷移する確率を示す。FはSの部分集合で最終状態の集合である。η^Fはn次元列ベクトルで、そのi番目の要素η_iは$s_i \in F$ならば$\eta_i = 1$，$s_i \notin F$ならば$\eta_i = 0$と定められる。

今 $x = \sigma_{i_1} \sigma_{i_2} \cdots \sigma_{i_k}$　$\sigma_{i_j} \in \{\varepsilon\} \cup \Sigma$ (εは空系列を示す)に対して、$A(x) = A(\sigma_{i_1}) \cdots A(\sigma_{i_k})$　(但し$A(\varepsilon)$は

マトン$\mathfrak{A}_0$が研究された。$\mathfrak{A}_0 = \langle S, \pi, \{A(0), A(1)\}, \eta^F \rangle$, $S = \{s_1, s_2\}$, $\pi = (1, 0)$ $A(0) = \begin{pmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$ $A(1) = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 1 \end{pmatrix}$ $\eta^F = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ とする。これによって受理される確率的事象を P とすると，$x = \sigma_{i_1} \sigma_{i_2} \cdots \sigma_{i_k}$ $\sigma_{i_j} \in \{0, 1\}$ に対して，$P(x) = 0.\sigma_{i_k} \sigma_{i_{k-1}} \cdots \sigma_{i_1}$ （2進展開）となることは容易にたしかめられる。$0 \leq \lambda < \lambda_1 \leq 1$ に対して，$T(\mathfrak{A}_0, \lambda)$ は $T(\mathfrak{A}_0, \lambda_1)$ を真に含むから，この有限確率オートマトンだけについても，ストカスティック言語は実数の濃度だけ異るものが存在し，したがってストカスティック言語の内にはチューリングマシンによって規定される言語の族に入らないものが存在することがわかる。そこで，ストカスティック言語でない言語はどのようなものかということが興味の対象となるがしばらくの間 Bukharaev (1965) が recursive set の範囲内に1ーシンボルからなるアルファベット上でその例を見出したと伝えられるにとどまっていた。しかし，コンテキスト・フリー言語の内にもストカスティックでない言語は存在するであろうと予想されていた。（例えば Salomaa (1969)），最近 Paz (1969) が非常に簡明な方法によって，1ーシンボルアルファベット上のコンテキスト・センシティブ言語の内にストカスティック言語でない言語の例を見出した。すなわち，$\Sigma = \{\sigma\}$ とし，a, b 2シ

ンボルからなる全ての系列を辞書式にならべて，さらにそれらを順番につないでできる無限系列Xにおいて，k番目のシンボルがσであるとき，かつそのときに限り，$\sigma^k \in L_s$とすることによってL_sを定めると，L_sはストカスティック言語ではない。このことを証明するまえに，確率的事象の族$\mathcal{P}$がfuzzy eventの族$\mathcal{F}$の真の部分集合であることを示す簡単な性質について述べておく。Pが確率的事象であると，任意の$x \in \Sigma^*$に対して，$c_0+c_1+\cdots+c_n=1$ （$c_n=1$）なる有限個の実数の集合$\{c_0, c_1, \cdots, c_n\}$が存在して，任意の$y, z \in \Sigma^*$に対して，$c_n P(yx^n z) + c_{n-1} P(yx^{n-1} z) + \cdots + c_0 P(yz) = 0$とすることができる。このことは，Pがある有限確率オートマトン$\mathfrak{A} = \langle S, \pi, \{A(\sigma) \mid \sigma \in \Sigma\}, \eta^F \rangle$によって$P(yx^i z) = \pi A(y) A(x)^i A(z) \eta^F$と書けることと，$A(x)$のcharacteristic polynomialを考えてみると容易にわかる。上に述べた性質をPazに従って，finiteness propertyということにする。

L_sがストカスティック言語であるとすると，適当な確率的事象Pとカットポイントλによって，$L_s = \{x \in \Sigma^* \mid P(x) > \lambda\}$とかける。finiteness propertyによって，任意のiに対して

$C_n + \cdots + C_0 = 0$，$C_n P(\sigma^{n+i}) + C_{n-1} P(\sigma^{n+i-1}) + \cdots + C_0 \cdot P(\sigma^i) = 0$ ——(1) なる実数の組$C = \{C_0, \cdots, C_n = 1\}$が存在する。Cの内の正数を$C_{j_1}, C_{j_2}, \cdots, C_{j_p}$とする。$L_s$の構成

法から明かなように，適当なiをとれば，$\sigma^{i}, \sigma^{i+1}, \cdots, \sigma^{i+n}$の内$\sigma^{j_1+i}, \sigma^{j_2+i}, \cdots, \sigma^{j_p+i}$は$L_s$の元で，その他は$L_s$に含まれないようにすることができる。$P(\sigma^{j+i}) > \lambda \Leftrightarrow \sigma^{j+i} \in L_s$ $(j=0,1,\cdots,n)$であるから，$C_n P(\sigma^{n+i}) + C_{n-1} P(\sigma^{n+i-1}) + \cdots + C_0 P(\sigma^{i}) = C_n(P(\sigma^{n+i}) - \lambda) + \cdots + C_0(P(\sigma^{i}) - \lambda) > 0$ となって(1)と矛盾が生じる。したがってL_sはストカスティック言語ではない。

同じ手法によって，ストカスティック言語ではないコンテキスト・フリー言語の存在を証明することができる。2-シンボルアルファベット$\{a, b\}$上で，$a^{i}ba^{j_1}b \cdots a^{j_r}b$ $(r=1,2,\cdots)$の形を持ち，適当なℓ $(1 \leq \ell \leq r)$に対して，$i = j_1 + j_2 + \cdots + j_\ell$が成立するようなテープの集合を$L_f$とする。なお$i, j_1, \cdots, j_r$は非負整数とする。そうすると$L_f$は線形コンテキストフリー文法で生成されることが示される。このことによって，コンテキスト・フリー言語の内でも，正規集合の直接的拡張とみられる簡単な言語の範囲内にストカスティック言語でないものが存在したことがわかった。後に述べるように，L_fはストカスティック言語の closure property に関する事実を証明するのにも有効である。

§3，closure property

有限確率オートマトンに関連して，2つのタイプの closure

propertyの問題がある。1つは確率的事象の族$\mathfrak{P}$のfuzzy setの意味での演算に関する問題であり，もう1つは，もっと広く，ストカスティック言語の族の演算に関して考える問題である。

$\mathfrak{P}$のclosure propertyに関して次の結果が得られる。

(1). $p, q, r \in \mathfrak{P}$ に対して，$(p, q; r)(x) = p(x)r(x) + q(x) \cdot (1 - r(x))$ $(x \in \Sigma^*)$ で与えられるfuzzy event $(p, q; r)$（コンベクス・コンビネーション）は確率的事象である。

(1)より(2)が導かれる。

(2)(和集合，共通集合)，$p, q \in \mathfrak{P}$ に対して，fuzzy event $p \vee q$，$p \wedge q$ を $p \vee q(x) = \max(p(x), q(x))$，$p \wedge q(x) = \min(p(x), q(x))$ $(x \in \Sigma^*)$ で定義すると，$p, q \in \mathfrak{P}$ でかつ $\{x \in \Sigma^* \mid p(x) > q(x)\}$ が正規集合ならば，$p \vee q$，$p \wedge q$ は共に確率的事象である。

上の性質は次のように表現を変えると，fuzzy setの概念を導入する意義がわかる。

(2′) 有限確率オートマトン$\mathfrak{A}_1$，$\mathfrak{A}_2$がそれぞれ，$p, q \in \mathfrak{P}$を受理するとき，$\{x \in \Sigma^* \mid p(x) > q(x)\}$が正規集合ならば，ある有限確率オートマトン$\mathfrak{A}_3$，$\mathfrak{A}_4$が存在して，任意のカットポイント $0 \leq \lambda \leq 1$ に対して，$T(\mathfrak{A}_3, \lambda) = T(\mathfrak{A}_1, \lambda) \cup T(\mathfrak{A}_2, \lambda)$

$T(\mathfrak{A}_4, \lambda) = T(\mathfrak{A}_1, \lambda) \cap T(\mathfrak{A}_2, \lambda)$ が成立する。

$p, q \in \mathfrak{P}$ に対して，$p \vee q$，$p \wedge q$ は一般的には$\mathfrak{P}$の元でな

いことは，finiteness property を用いて，松浦が最初に証明した。

(3)（逆事象）$p \in \mathfrak{P}$，$p^T(x) \underset{\mathrm{def}}{\equiv} p(x^T)$ $(x \in \Sigma^*)$，但し x^T はテープの逆にならべたものを示す。このとき $p^T \in \mathfrak{P}$。

(3') 任意の有限確率オートマトン $\mathfrak{A}$ に対して，ある有限確率オートマトン $\mathfrak{A}_1$ が存在して，任意のカットポイント $0 \le \lambda \le 1$ に対して，$T(\mathfrak{A}, \lambda)^T = T(\mathfrak{A}_1, \lambda)$ が成立する。

(4) Σ，Δ をそれぞれアルファベットとする。Σ^* から Δ^* への gsm 写象 Γ と Δ^* 上の確率的事象 p に対して，$\hat{\Gamma}(p)(x) = p(\Gamma(x))$ $(x \in \Sigma^*)$ で与えられる Σ^* 上の fuzzy event $\hat{\Gamma}(p)$ は確率的事象である。

(4') Γ を Σ^* から Δ^* への gsm 写像とする，Δ 上の任意の有限確率オートマトン $\mathfrak{A}_\Delta$ に対して，ある Σ 上の有限確率オートマトン $\mathfrak{A}_\Sigma$ が存在して，任意のカットポイント $0 \le \lambda \le 1$ に対して $\Gamma^{-1}(T(\mathfrak{A}_\Delta, \lambda)) = T(\mathfrak{A}_\Sigma, \lambda)$ が成立する。

有限確率オートマトン $\mathfrak{A} = \langle S, \pi, \{A(\sigma) \mid \sigma \in \Sigma\}, \eta^F \rangle$ において，$A(\sigma), \sigma \in \Sigma$，$\pi$ の各要素が全て有理数とするとき，$\mathfrak{A}$ を有理数有限確率オートマトン（RFPA）という。RFPA によって受理される確率的事象を有理数確率的事象（RPE）ということにし，その族を $\mathfrak{P}_R$ とする。そうすると $p, q \in \mathfrak{P}_R$ に対して，$\{x \in \Sigma^* \mid p(x) = q(x)\}$ 及び $\{x \in \Sigma^* \mid p(x) \neq q(x)\}$

はストカスティック言語となる。（松浦,等 . Turakainen 1968）

$\{x \in \Sigma^* \mid p(x) = q(x) \quad p, q \in \mathfrak{P}_R\}$ なる形の言語をE-言語,
$\{x \in \Sigma^* \mid p(x) \neq q(x) \quad p, q \in \mathfrak{P}_R\}$ なる形の言語をD-言語,
$\{x \in \Sigma^* \mid p(x) > q(x) \quad p, q \in \mathfrak{P}_R\}$ なる形の言語をP-言語とそれぞれいうことにすれば，P言語はコンテキスト・センシティブ言語に真に含まれる（都倉,等 1968）。E言語でもD言語でもないP言語は存在する（例えば $\{a^m b^n \mid m \geq 0 \ n \geq 0 \ m > n\}$）けれども，E言語とD言語はもともと具体的に作りやすいストカスティック言語である。前の性質(4)を用いて，任意のΣ^*からΔ^*へのgsm写像Γ_1, Γ_2に対して，$\{x \in \Sigma^* \mid \Gamma_1(x) = \Gamma_2(x)\}$，$\{x \in \Sigma^* \mid \Gamma_1(x) \neq \Gamma_2(x)\}$なる形の言語はストカスティック言語であることがわかり，又同じ考え方により，次のような

規則Pを持つ文法(V, Σ, P, σ)（d.l文法という）で生成される言語（d.l言語という）はストカスティック言語であることがいえる。(i) $v \to a\xi u$ あるいは $v \to b$, $v, \xi \in V - \Sigma$, $a, b \in \Sigma$, $u \in \Sigma^*$ の形を持ち (ii) 任意の2つの規則 $v_1 \to a_1 x$, $v_2 \to a_2 y$, $a_1, a_2 \in \Sigma$, $x, y \in \{\varepsilon\} \cup (V - \Sigma)\Sigma^*$, $v_1, v_2 \in V - \Sigma$ に対して，$v_1 = v_2$ かつ $a_1 = a_2$ ならば $x = y$ がなりたつ。

$L_1 = \{a^i b a^{j_1} \cdots b a^{j_r} \mid i, j_1, \cdots, j_r \geq 0 \quad i = j_1 + \cdots + j_r\}$ はE言語したがってストカスティック言語である（Turakainen）。正規集合 $b(\varepsilon \cup \{a, b\}^* b)$ を L_2 とすると，$L_1 L_2 = L_f$ となる

から，ストカスティック言語の族は　コンカテネーションの演算のもとで閉じていない。又 $L_3 = L_1 c(\Sigma \cup \{a,b\}^* b)$ もストカスティック言語であることが証明できる。そこで $h(a)=a$, $h(b)=b$ $h(c)=b$ なるホモモルフィズムを考えると，$h(L_3)=L_f$ となるから，ストカスティック言語の族はホモモルフィズムの演算のもとで閉じていない。Turakainen は L_f をあたえられて，我々と独立に上の結果を得，　　さらに $L'=\{a^k b (a^* b)^* a^k b \mid k \geq 0\}$ がストカスティック言語であることを示し，L'^* がストカスティック言語でないことを同様の手法で証明した。

§3. いくつかの決定不可能問題

最後にRFPAに関する二，三の決定問題について述べる。以下の事実は，E-言語によって，ストカスティック言語を構成し，Post の Correspondence problem に帰着させることによって証明できる。

(1) 任意のRFPA $\mathfrak{A}$ と有理数カットポイント λ に対して，イ $T(\mathfrak{A},\lambda)$ は空か否か，ロ，$T(\mathfrak{A},\lambda)$ は Σ^* か否か　ハ，$T(\mathfrak{A},\lambda)$ は正規集合であるかどうか　ニ，$T(\mathfrak{A},\lambda)$ はコンテキスト・フリーであるか　の4つの問題は全て帰納的に決定不可能である。

(2) 任意のRFPA $\mathfrak{A}_1$ と $\mathfrak{A}_2$ に対して，それが受理する確率的事象をそれぞれ P_1, P_2 とすると，イ，$P_1 \vee P_2$ は確率的事象で

あるか ロ，$P_1 \wedge P_2$ は確率的事象であるか の２つの問題は帰納的に決定不可能である。

参考文献

Bar-Hillel, Y., Perles, M. and Shamir, E.: On formal properties of simple phrase structure grammars, Z. Phonetic, Sprach, Kommu-nicationsforsch., vol. 14 (1961) 143-

Bukharaev, R.: Criteria for the representation of events in finite probabilistic automata. Dokl. Akad. Nauk SSSR 164 (1965)

Ginsburg, S.: Mathematical Theory of Context Free Languages. McGraw-Hill, New York (1966)

Harrison, M. A.: Introduction to Switching and Automata Theory McGraw-Hill, New York (1965)

松浦，稲垣，福村：オートマトンの一般化とその解析 オートマトン研究会資料 (1968-1)

松浦，稲垣；福村：線形空間オートマトンと確率オートマトン．電子通信学会全国大会講演論文集 58-4 昭和43年10月

Nasu, M. and Honda, N. Fuzzy events realized by finite probabilistic automaton Information and Control 12 284-303 (1968)

Nasu, M and Honda, N. Mappings induced by PGSM-mappings and some recursively unsolvable problems of finite probabilistic automata. Information and Control 15 250-272 (1969)

Nasu, M. and Honda, N. A context free language which is not acceptable by a probabilistic automaton. Information and Control 掲載決定 (1971)

Paz, A. Some aspects of probabilistic automata, Information and Control 9 26-60 (1966)

Paz, A. Events which are not representable by a probabilistic automaton (preliminary draft) (1969)

Paz, A. Formal Series, Finiteness properties and Decision Problems Technical Report NO.4 ISRAEL INSTITUTE OF TECHNOLOGY (1970)

RAbin, M.O. Probabilistic Automata, Information and Control 6 230-245 (1963)

Salomaa, A. Theory of Automata Pergamon Press Oxford (1969)

Salomaa, A., Probabilistic and weighted grammars. Information and Control 15 529-544 (1969)

Schützenberger, M.P. Certain elementary families of automata, Symposium on Mathematical theory of Automata, Polytechnic Institute of Brooklyn (1962)

都倉, 藤井, 嵩: 線形オートマトンに関する二,三の考察 電子通信学会全国大会講演論文集 S8-2 昭和43年10月

Turakainen, P. On probabilistic automata and their genera

-lizations. ANNALS ACADEMIE SCIENTIARUM FENICAE Series A I MATHEMATICA 424 (1968)

Turakainen, P., The family of stochastic languages is cloed neither under catenation nor under homomorphism. Ann. Univ. Turku, A I 133 (1970)

Turakainen, P., Some closure properties of the family of stochastic languages, manuscript, (1970)

Zadeh, A.L., Fuzzy sets. Information and Control 8 338-353 (1965)